

Invasive Species Early Detection Protocol for Klamath Network Parks

Standard Operating Procedure (SOP) #11: Metadata Guidelines

Version 1.00 (February 2010)

Revision History Log:

Previous Version	Revision Date	Author	Changes Made	Reason for Change	New Version

Acronyms used in this SOP are defined at the end.

This SOP explains the procedures for completing metadata for products developed using this protocol. This includes, but is not limited to, databases, documents, GPS data, and GIS data. Details on metadata for photographs can be found in SOP #7: Photo Management. This SOP is based on metadata recommendations developed by the Klamath Network (KLMN) (Mohren 2007), the Natural Resource GIS Program (NR-GIS Data Store 2005a-i), and the NPS North Coast Cascades Network (NCCN 2006a, NCCN 2006b, NCCN 2006c).

Documentation is a critical step towards ensuring that products collected as part of this monitoring effort are usable for their intended purposes now and in the future. This involves the development of metadata, which is defined as structured information about the content, quality, condition, and other characteristics of a product. In addition to spatial information, metadata includes information about data format, collection and analysis methods, time of collection, originator, access/use constraints, and distribution. Metadata provide the means to catalog products, within Intranet and Internet systems, making them available to a broad range of potential users. While most frequently developed for geospatial data, metadata describing non-geospatial datasets are also important (NCCN 2006a).

Timelines

It is the responsibility of the Crew Lead to submit metadata or metadata products (e.g., Metadata Interview Form, data dictionary) to the Klamath Network Data Manager, in the proper format, when he or she submits the product with which the metadata are associated. SOP #8: Data Transfer, Storage, and Archive provides the details on the products to be delivered to the Data Manager, the due dates for those products, and the persons responsible for those products.

Responsibilities and Standards

Metadata are among the most important pieces of documentation to help guarantee the long-term usability of data. The degree of documentation will vary depending on the product, but a few standards will always hold true.

1. Metadata for spatial data collected through I&M funded projects will meet FGDC, NBII, and NPS standards.
2. Crew Leads will be expected to submit a data dictionary (for tabular and spatial data) and a Metadata Interview Form (for spatial data) prior to the start of the first field season.
3. Crew Leads will be expected to review and revise all data dictionaries and Metadata Interview Forms at the end of each field season and report changes following the timeline listed in SOP #8: Data Transfer, Storage, and Archive.
4. It is the responsibility of the Data Manager to develop the official metadata based on the data dictionary and Metadata Interview Form provided by the Crew Lead.
5. It is the Data Manager's responsibility to parse and transfer metadata to the NPS Data Store, if applicable.
6. The Data Manager will work with the Project Lead and park staff to determine the sensitivity level of any data.

Reports

Three main types of reports are expected to be developed during this monitoring effort, including Biennial Reports, Analysis and Synthesis Reports, and Scientific Publications (SOP #10: Reporting and Analyses of Data). It is the responsibility of the individual creating the reports to ensure that the following guidelines are met.

1. First and last name of all authors are included on the report.
2. Affiliations of the authors are included on the report.
3. Version numbers are used on all drafts of the report.
4. The date the report was completed is included on the report.
5. The date representing the information presented in the report is included in the report.
6. Series number is included in the report when applicable.
7. The NatureBib accession number has been added to the subject field in the properties of the document in the format: NatureBib #123456. This will be created by the Network Data Manager.

Spatial Data, Databases, and Spreadsheets

The Data Manager, working closely with the GIS Specialist, is responsible for creating and maintaining the official metadata for all GIS and GPS products, relational databases, and spreadsheets. It is the responsibility of the Crew Lead to provide Metadata Interview Forms and data dictionaries to the Data Manager prior to implementing the field work as described in SOP #8: Data Transfer, Storage, and Archive. The Data Manager will use the Metadata Interview Form, data dictionary, and protocol to develop complete metadata for each product. Each year that field work occurs, the Crew Lead must submit a data dictionary (if changes to the product have occurred) and Metadata Interview Form following the timeline in SOP #8: Data Transfer, Storage, and Archive.

Steps for Metadata Creation

Step 1: Metadata Interview Form and Information Gathering (Figure 1)

- A. The Crew Lead should obtain and complete the KLMN Metadata Interview Form at project onset to facilitate compiling the information required to create compliant metadata.
 - 1. The KLMN Metadata Interview Form is posted on the KLMN Internet and Intranet web pages. In addition, the form can be obtained by contacting the Network Data Manager.
 - 2. Best attempts should be made to populate the Metadata Interview Form as completely as possible prior to starting field work. However, it is recognized that changes to the form will occur throughout the project.
- B. A data dictionary must be created to provide information to help the Data Manager create or update the official metadata and in some cases, the project database. The data dictionary should include:
 - 1. The name and purpose of each table, shapefile, coverage, or feature class.
 - 2. A list of attribute names, types, sizes, and descriptions by table. An example of a data dictionary in the proper format is provided at the end of this SOP.
 - 3. If the database was not designed by the KLMN, it is the responsibility of the Project Lead to provide the Data Manager with the data dictionary.
- C. If a taxa list other than a current ITIS certified taxa list was used, the Crew Lead will need to provide the list that was used at the end of each field season. Taxa lists should include:
 - 1. Taxon group (plant).
 - 2. Scientific name.
 - 3. Common name.
 - 4. Any special code that defines a species.
- D. The Crew Lead should send a copy of any additional information that might be valuable for the development of metadata.

Step2: Sensitivity Review: Sensitive Data (Species Locations, Site Locations, etc.) May not be Subject for Release to the Public

- A. The current version of NPS Data Store does not screen for sensitive information. Therefore, any data with a sensitive status will not be posted on the Data Store.
- B. The Network Data Manager will be responsible for posting data as sensitive. Status of the data will be based on comments provided by the Crew Lead under the “Sensitivity” question in the Metadata Interview Form and certification form. In addition, the Klamath Network will consult with park staff if the sensitivity status of any data is questionable.

Step 3: Metadata Software Selection

- A. The Klamath Network will utilize ArcCatalog, NPS Metadata Tools and Editor, and the Database Metadata Extractor to create metadata for all projects.
 - 1. ArcCatalog automatically harvests spatial organization and reference information, as well as entity and attribute information for GIS datasets.

2. The NPS Metadata Tools and Editor is provided as a standalone program or as an extension for ArcCatalog and is available at:
<http://science.nature.nps.gov/nrdata/tools/>
 - i. It can be used for metadata creation and editing.
 - ii. It can import, export, and parse metadata.
 - iii. It cannot harvest entity and attribute information; however, this is an anticipated feature for the next version.
3. The NPS Database Metadata Extractor (MTE) is a custom software application for authoring, editing, and managing NPS metadata. The MTE operates either as an extension to ArcCatalog versions 8.3/9.x or as a standalone desktop application. Eventually, this tool will be incorporated into the NPS Metadata Tools and Editor. Features of this tool include:
 - i. Automatically harvests entity (table) and attribute (field) metadata from Access databases, including domains.
 - ii. Allows the user to edit and review the harvested metadata and make batch edits.
 - iii. Allows the user to export metadata to an FGDC-compliant XML file.
 - iv. Allows exported XML to be used in the Metadata Tools and Editor either by opening it to start a new metadata record or by updating it with a template to fill in Section 5 of an existing metadata record.

Step 4: Additional Requirements

- A. Along with the required metadata, the Klamath Network requires the following information be included in the metadata document.
 1. The name and agreement code for the project. These references can be entered in the Related Key element in the Program Information section (NPS Section 0) on the NPS Profile.
 2. References to all products (GIS, GPS, Databases, Reports) generated by the projects. These references can be entered in the repeatable Cross Reference element of the Identification Information section.
 3. Standard language for NPS liability should be inserted into the Distribution Liability metadata element of the Distribution Information (FGDC Section 6). This can be found at: <http://www.nps.gov/gis/liability.htm>.

Step 5: Biological Data Profile

If a dataset includes biological information, the Biological Data Profile provides a set of extended metadata elements to document the species observed, taxonomic information, methods, and analytical tools.

- A. The most direct, and KLMN preferred, means to populate the Biological Data Profile metadata elements are outlined in [Biological Profile \(National Biological Information Infrastructure - NBII\) Metadata Guide](#) (NR-GIS Data Store 2005a).
 - i. This approach primarily utilizes the NPS Metadata Tools and Editor and may also require the entity and attribute harvesting capability of NPS Database Metadata Extractor for Access datasets.
- B. The two documents at the following links describe alternative approaches to completing the Biological Data profile for a metadata record. Note that the first

requires the use of additional metadata creation software (Spatial Metadata Management System, or SMMS):

- i. [Metadata Tools Used in the Creation of the FGDC Biological Data Profile](#) (Callahan and Devine 2004).
- ii. [National Biological Information Infrastructure \(NBII\) Metadata Steps](#) (McGuire 2004).

Step 6: Metadata Review

The Data Manager should review metadata for quality control (QC) prior to posting to NPS Data Store. A useful QC Checklist is available for download on the NPS Intermountain Region GIS website at: http://imgis.nps.gov/tips_templates.html.

Step 7: Metadata Parsing and Exporting to XML Format

The NPS Data Store requires that metadata records be parsed into FGDC-structured metadata and then exported to XML format.

- A. If using ArcCatalog, these steps can both be done directly with the NPS Metadata Tools and Editor. See [Parsing Metadata with the NPS Metadata Tools and Editor](#) (NR-GIS Data Store 2005i) for more information.
- B. If using other applications, export the metadata first to ASCII text format and then parse with Metadata Parser (MP). MP can simultaneously output an XML format metadata file as well.
 - i. MP must be customized to handle NPS, Biological Data, or ESRI Profile metadata elements. For specifics, refer to:
 - a. The README.txt file included in the zipped NPS Metadata Profile configuration files, available from the NPS Data Store website at: <http://science.nature.nps.gov/nrdata/docs/metahelp/metahelp.cfm>.
 - b. [Parsing Metadata with the NPS Metadata Tools and Editor](#) (NR-GIS Data Store 2005i).

Step 8: Metadata Posting

Post the metadata to the NPS Data Store.

- A. Authorized NPS staff may request upload and edit access to the NPS Data Store through the NPS Natural Resource Universal Web Login (UWL), available at: <https://science1.nature.nps.gov/nrdata/>. This is also the portal for uploading data.
- B. More information about metadata upload format requirements is available at: <http://science.nature.nps.gov/nrdata/docs/metahelp/metainfo.cfm> and in [Metadata and Data Uploading Guidance](#) (NR-GIS Data Store 2005g).

Step 9: Editing/Updating Metadata Already Posted to NPS Data Store

As of Version 1, the NPS Data Store application allows online editing of NPS Theme Category and ISO Theme Keyword information and the deletion of single metadata records and/or datasets only (see help documentation at:

<http://science.nature.nps.gov/nrdata/docs/metahelp/edithelp.cfm>).

- A. For metadata records simply needing edits to NPS Theme Category or ISO Theme Keyword elements, refer to [Editing Category Information](#) (NR-GIS Data Store 2005f).

- B. If a metadata record posted to the NPS Data Store contains errors or requires edits to other elements, it will need to be deleted from the NPS Data Store, edited, and then reposted. Refer to [Deleting Single Records](#) (NR-GIS Data Store 2005e).
- i. The user should first download the metadata record (save in XML format) to the local system, then edit as needed in a text editor or metadata software program.
 - ii. The edited metadata record can then be resubmitted to the NPS Data Store.
 - iii. If the dataset documented by the metadata record requires no edits, it will not need to be reposted. Simply ascertain that the metadata file still specifies the correct pathway to the dataset on the NR-GIS Data Server before resubmitting the metadata file.

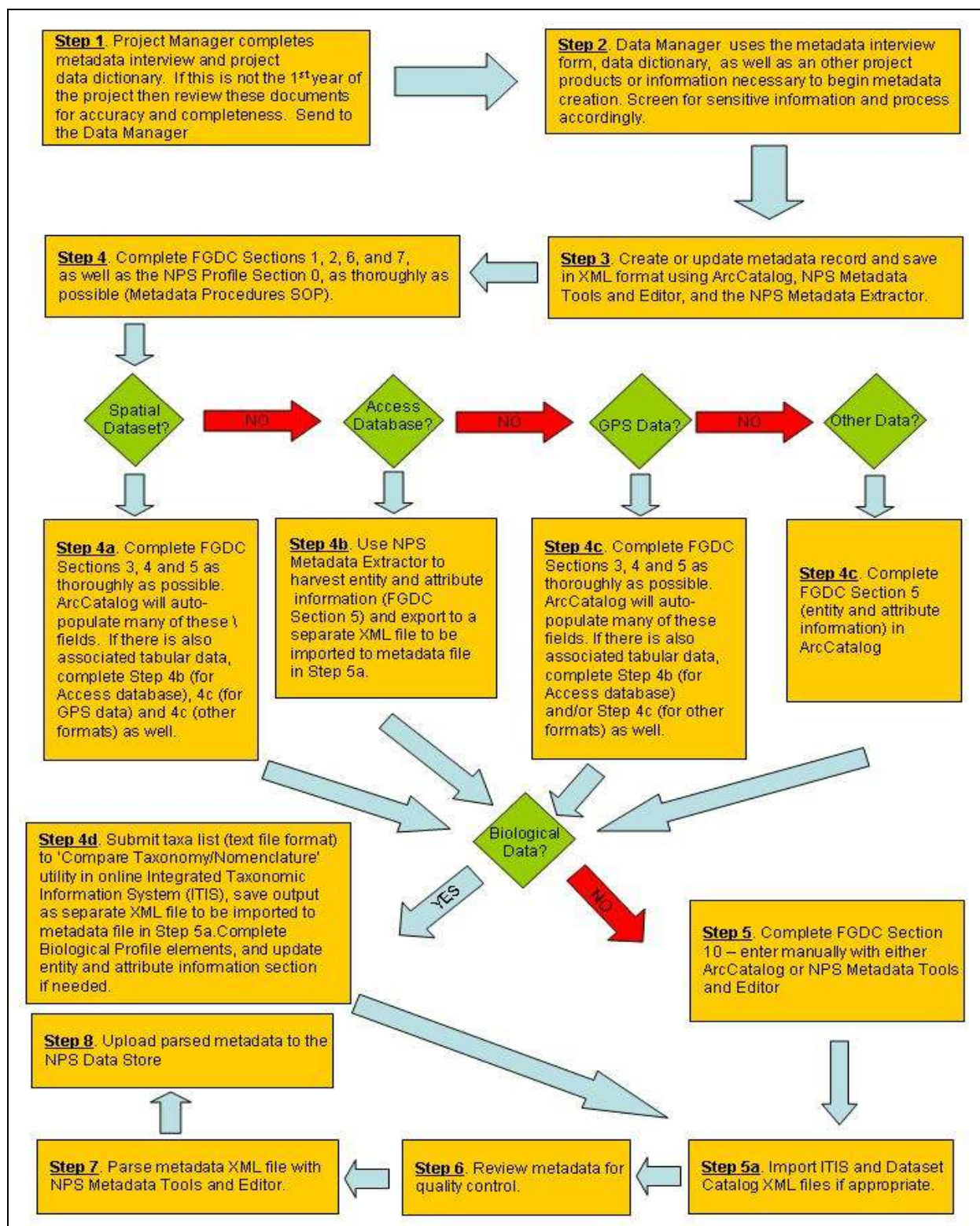


Figure 1. The general workflow for metadata creation for spatial and tabular data.

Literature Cited

- Callahan, K., and H. Devine. 2004 (draft). Metadata tools used in the creation of the FGDC biological data profile. National Park Service, Northeast Region. Online. (http://science.nature.nps.gov/im/datamgmt/docs/SOP_BioDataProfileTools_v1.doc). Accessed 6 March 2008.
- McGuire, S. 2004. National Biological Information Infrastructure (NBII) metadata steps. National Park Service, Midwest Region. Online. (<http://www.nature.nps.gov/im/units/mwr/gis/download/BiologicalProfileGuide.pdf>). Accessed 6 March 2008.
- Mohren, S. R. 2007. Data management plan, Klamath Inventory and Monitoring Network. Natural Resource Report NPS/KLMN/NRR—2007/012. National Park Service, Fort Collins, Colorado.
- North Coast and Cascades Network – National Park Service. 2006a. Metadata development guidelines. National Park Service. Online. (http://science.nature.nps.gov/im/units/nccn/datamgmt_guide.cfm). Accessed 1 June 2008.
- North Coast and Cascades Network – National Park Service. 2006b. Metadata procedures SOP. National Park Service. Online. (http://science.nature.nps.gov/im/units/nccn/datamgmt_guide.cfm). Accessed 1 June 2008.
- North Coast and Cascades Network – National Park Service. 2006c. NCCN metadata interview. National Park Service. Online. (http://science.nature.nps.gov/im/units/nccn/datamgmt_guide.cfm). Accessed 1 June 2008.
- NR-GIS Data Store. 2005a. Biological profile (National Biological Information Infrastructure - NBII) metadata guide. National Park Service, Natural Resource and GIS Programs. Online. (<http://science.nature.nps.gov/nrdata/docs/metahelp/BiologicalProfileGuide.pdf>). Accessed 3 March 2008.
- NR-GIS Data Store. 2005b. Creating metadata. National Park Service, Natural Resource and GIS Programs. Online. (<http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISCreatingMetadata.pdf>). Accessed 3 March 2008.
- NR-GIS Data Store. 2005c. Creating non-geospatial metadata. National Park Service, Natural Resource and GIS Programs. Online. (<http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISDataStoreCreatingNGSMetadata.pdf>). Accessed 3 March 2008.
- NR-GIS Data Store. 2005d. Creating simple geospatial metadata. National Park Service, Natural Resource and GIS Programs. Online. (<http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISDataStoreCreatingGSMetadata.pdf>). Accessed 3 March 2008.
- NR-GIS Data Store. 2005e. Deleting single records. National Park Service, Natural Resource and GIS Programs. Online. (<http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISDataStoreDeleting.pdf>). Accessed 3 March 2008.

- NR-GIS Data Store. 2005f. Editing category information. National Park Service, Natural Resource and GIS Programs. Online.
(<http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISDataStoreEditCategory.pdf>). Accessed 3 March 2008.
- NR-GIS Data Store. 2005g. Metadata and data uploading guidance. National Park Service, Natural Resource and GIS Programs. Online.
(<http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISMetadataDataUploadGuidance.pdf>). Accessed 3 March 2008.
- NR-GIS Data Store. 2005h. Metadata authoring guidance. National Park Service, Natural Resource and GIS Programs. Online.
(<http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISMetadataAuthoringGuidance.pdf>). Accessed 3 March 2008.
- NR-GIS Data Store. 2005i. Parsing metadata with the NPS Metadata Tools and Editor. National Park Service, Natural Resource and GIS Programs. Online.
(<http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISMetadataParsingGuidance.pdf>). Accessed 3 March 2008.

Electronic Metadata Interview

[Note: Please make your responses directly within this word document in "Red" type.]

1. Have you already prepared metadata for this dataset?
 - a. If yes, please send a copy of the document or reference to where it can be found and skip to item **18**.
2. What is the title of the dataset?
3. Who is the originator(s)/owner of the dataset? (Include address and telephone number)?
 - a. If someone else should answer question about the data, please list the name, address, and telephone number.
 - b. Are there other organizations or individuals who should get credit for support, funding, or data collection and analysis?
4. Does the dataset contain any sensitive information that should not be released to the public? NPS?
 - a. Explain why the data should not be released to the public.
 - b. Explain why the data should not be released to non-park NPS staff.
5. Is the dataset published or part of a larger publication?
 - a. If so, what is the reference?
6. Include a brief (no more than a few sentences) description of the dataset.
7. Why were the data collected in the first place?
8. What is the time period represented by the dataset?
9. Were the data developed primarily through:
 - a. Field visits?
 - b. Remote instrumentation (e.g., temperature recorders, etc.)?
 - c. Existing data sources?
10. What is the status of the data you are documenting? – *complete, in progress, planned*
 - a. Will the dataset be updated? If so, how frequently?
11. Where were the data collected? Include description and coordinates, if known.
12. List some keywords to help search for this dataset.
 - a. Thematic, Place, Temporal, Strata, Taxonomy
 - b. If a controlled vocabulary was used, what is the reference?

13. List any related datasets that could be documented for cross-reference.
14. The FGDC Biological Profile includes the means to document tabular datasets, taxonomy, field methods, and the use of analytical tools or models.
 - a. Was your dataset developed using a model or other analytical tool?
 - i. If so, what is the reference?
 - ii. If the model or tool is available, include a contact and/or URL.
 - b. Does the dataset contain biological information? If no, skip to item **15**.
 - i. What species or communities were examined?
 - ii. Did you use a taxonomic authority or field guide for identification? If so, what is the reference?
 - iii. Briefly summarize your field methods (cut and paste from other documents!).
 - iv. If you used existing protocols or methods, list the references.
 - v. If you use a different taxonomic hierarchy than what is available in ITIS, then you need to supply the taxonomic hierarchy for all species within the dataset.
15. Is your dataset archived in a databank or data catalog? If yes, please include a reference to the documentation and skip to item 16. If No:
 - a. What measures did you take to make certain that your dataset was as nearly correct as possible?
 - b. Were there any things that you excluded from your data collection (e.g., stems less than a certain diameter or streams without surface flow)?
 - c. What is the form of your dataset? - *spreadsheet, ASCII file, GIS layer, database, other*.
 - d. What is the filename for your dataset?
 - i. For each file or table, list the fields in the dataset and for each field list:
 - The definition of the field.
 - If the data are coded (Enumerated Domain), list the codes and the definitions.
 - If the codes come from a published code set (Codeset Domain), list the reference.
 - If the data are measured (Range Domain), list the units and the minimum and maximum allowable values ("no limit" is acceptable).

- ii. Otherwise, the domain is unrepresentable. Include a brief description of what is in the field.
- 16. Is this a GIS dataset? If no, skip to item 17.
 - a. Include a path to where the data can be accessed over the network or send a copy of the ArcInfo export file, an ArcView shapefile, or an ArcCatalog exported metadata file (txt or xml).
 - i. Include projection parameters, if necessary.
 - b. List any source datasets you used. For each source, list:
 - i. Source name, originator, and publication date.
 - ii. Source time period and scale.
 - iii. Source presentation form and media type.
 - iv. Contribution of source to your analysis.
 - c. List the processing steps you used to create your dataset, including the approximate date of processing.
- 17. Is the dataset available for distribution? If no, go to 18.
 - a. Are there legal restrictions on who may use the data?
 - b. Do you have any advice for potential users of the dataset?
 - c. What are your distribution instructions?
- 18. You are done. Send this completed document with the relevant responses to this interview to your metadata coordinator (Sean Mohren, Klamath Network Data Manager Sean_Mohren@nps.gov, 541-552-8576).

Example Data Dictionary

Dataset: ASXXYY

File Type: mdb

Relationship: Area Search Header (one) to Area Search Data (many)

Table: Area Search Header

Field Name	Field	Required (Y/N)	Type	Length	Decimal	Definition	Enumerated Domain	Range Domain
RecNum	1	Y	Numeric	Integer	0	Auto Number based on order of entry; Key Field		Integer starting at 1, no limit
Project	2	Y	Character	20		Project or Region Code		
Site	3	Y	Character	20		Site name (often 4-letter code)	Each 4-letter code represents itself	
Point	4	Y	Character	2		Point (Search area)	Typically designated A or B	
Month	5	Y	Character	2		Month		01 to 12
Day	6	Y	Character	2		Day		01 to 31
Year	7	Y	Character	4		Year		
PrimObs	8	Y	Character	4		Primary observer's initials		
SecObs	9	N	Character	20		Secondary observer's initials, if multiple observers, initials separated by “ , ”		
Temp	10	Y	Character	3		Temperature		Degrees Celsius, range=-10 to 40
CldCvr	11	Y	Character	3		Cloud Cover		Percentage, range=0 to 100%

Definitions and Acronyms

<i>ArcCatalog</i>	Module in ESRI's ArcGIS software within which metadata for spatial datasets (coverages, shapefiles) can be created.
<i>Biological Data Profile</i>	Set of definitions for the documentation of biological data through the creation of extended elements to the FGDC Content Standard for Digital Geospatial Metadata (CSDGM).
<i>CSDGM</i>	Content Standard for Digital Geospatial Metadata. The FGDC-promulgated metadata standard established to provide a common set of terminology and definitions for documenting digital geospatial data.
<i>Dataset Catalog</i>	NPS Inventory and Monitoring Program tool for metadata creation, ideal for abbreviated dataset documentation but not for fully FGDC-compliant metadata creation.
<i>ESRI®</i>	Environmental Systems Research Institute. A GIS software company.
<i>FGDC</i>	Federal Geographic Data Committee. The interagency committee that promotes the coordinated development, use, sharing, and dissemination of geographic data.
<i>GIS</i>	Geographic Information System. A computer system for capturing, manipulating, analyzing, and displaying data related to positions on the Earth's surface.
<i>ISO</i>	International Organization for Standardization. A network of national standards institutes of 150 countries, responsible for the "ISO 19115" international metadata standard.
<i>Metadata</i>	Data about the content, quality, condition, and other characteristics of a dataset, documented in a standardized format.
<i>MP</i>	Metadata Parser. A command-line program developed by the USGS to locate syntax errors in metadata files, verify FGDC-compliance, and convert between file formats.
<i>NBII</i>	National Biological Information Infrastructure. Collaborative program instrumental in developing the Biological Data Profile of the FGDC's CSDGM.

NPS Profile

The NPS Natural Resource and GIS Metadata Profile extends the FGDC CSDGM to incorporate NPS-specific elements such as park and project details. The NPS Profile includes the Biological Data Profile and the ESRI Profile.

NPS Data Store

The NPS Natural Resource, GIS, and I&M Programs' web-based system (incorporating a database, data server, and secure web interface) to integrate data dissemination and metadata maintenance for Natural Resource, GIS, and other program data sets, digital documents, and digital photos.

SGML

Standard Generalized Markup Language. An ISO standard flexible markup language (predecessor to XML) used in many applications, including electronic publishing on the Web.

XML

Extensible Markup Language. A simple and flexible text format (a profile, or subset, of SGML) that facilitates large-scale electronic publishing and exchange of data on the Web.